**3GPP TSG-SA WG4 MBS SWG Meeting *S4aI240161***

**Online, 16th October 2024 - 18th October 2024**

|  |
| --- |
| *CR-Form-v12.3* |
| **CHANGE REQUEST** |
|  |
|  | **26.804** | **CR** | **<CR#>** | **rev** | **<Rev#>** | **Current version:** | **<Version#>** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network |  |

|  |
| --- |
|  |
| ***Title:***  | [FS\_AMD] WT 13 Updates from literature on HTTP3 |
|  |  |
| ***Source to WG:*** | Xiaomi |
| ***Source to TSG:*** | S4 |
|  |  |
| ***Work item code:*** | FS\_AMD |  | ***Date:*** | 14/10/2024 |
|  |  |  |  |  |
| ***Category:*** | **B** |  | ***Release:*** | Rel-19 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19) Rel-20 (Release 20)* |
|  |  |
| ***Reason for change:*** | Within FS\_AMD QUIC for segmented media streaming is studied. Within this study HTTP/3 delivery is documented in the Clause 5.4 of TR 26.804. This contribution updates this clause with insights from literature study. |
|  |  |
| ***Summary of change:*** | The following changes are proposed:* Update reference to add literature sources
* Add sources of potential issues of streaming clients operating on top of HTTP/3
* Add clause 5.4.5.8 detailing potential issues originating from processing overhead on the UEs when using HTTP/3
 |
|  |  |
| ***Consequences if not approved:*** | Clause 5.4 will be outdated. Potential conclusions of the study will not be complete. |
|  |  |
| ***Clauses affected:*** | 5.4 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  |  |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  |  |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  |  |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

Change #1

[104] IETF RFC 3986: "Uniform Resource Identifier (URI): Generic Syntax".

[x1] Chaudhary, Sapna, Mukulika Maity, Sandip Chakraborty, and Naval Shukla. "A dataset for analyzing streaming media performance over HTTP/3 browsers." *Advances in Neural Information Processing Systems* 36 (2024)

[x2] Chaudhary, Sapna, et al. "Managing Connections by QUIC-TCP Racing: A First Look of Streaming Media Performance Over Popular HTTP/3 Browsers." *IEEE Transactions on Network and Service Management* (2024).

[x3] Kazuho Oku, Lucas Pardue Robin Marx, Luca Niccolini, Marten Seemann, draft-kazuho-httpbis-http3-on-streams-00, "HTTP/3 on Streams", Expired Internet-Draft, 19 August 2024.

[x4] Chellappa, Sindhu, and Radim Bartos. "Is QUIC Quicker with HTTP/3? An Empirical Analysis of Quality of Experience with DASH Video Streaming." *2022 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE, 2022.

[x5] Herbots, Joris, et al. "Cross that boundary: Investigating the feasibility of cross-layer information sharing for enhancing ABR decision logic over QUIC." *Proceedings of the 33rd Workshop on Network and Operating System Support for Digital Audio and Video*. 2023.

[x6] Zhang, X., Jin, S., He, Y., Hassan, A., Mao, Z. M., Qian, F., & Zhang, Z. L. (2024, May). QUIC is not Quick Enough over Fast Internet. In *Proceedings of the ACM on Web Conference 2024* (pp. 2713-2722).

Change #2

### 5.4.5 Potential open issues

#### 5.4.5.1 Introduction

Most of the potential questions previously identified in clause 5.4.1 don’t appear to be continuing concerns, as HTTP/3 has been deployed on the Internet at scale. A few open issues are worth noting here.

#### 5.4.5.2 Streaming Protocols taking advantage of HTTP/3 capabilities

Because so many media providers have used HTTP-based transport mechanisms, they have been forced to provide “workarounds” to overcome the TCP-related deficiencies that HTTP implementations were forced to deal with. Confirming with the predominant implementors of HLS, MPEG-DASH, SRT, and even RTP what their timeframes are for versions of these protocols that take advantage of HTTP/3 capabilities would be useful, but these versions are already starting to appear. Preliminary proposals for “Tunnelling SRT over QUIC” [SRT-QUIC], and multiple proposals for RTP over HTTP/3, or even directly over QUIC, have been submitted to the IETF.

#### 5.4.5.3 3GPP-specific impediments to HTTP/3 deployment

One well-recognized impediment to HTTP/3 deployment in the broader Internet is that it is carried over QUIC, which is a well-behaved transport protocol that detects and responds to path congestion, but QUIC runs over the UDP protocol, which is not understood to be well-behaved. Traditional UDP protocols have been query–response protocols, such as DNS domain name resolution ("what is the IP address for this domain name?" followed by "This is the IP address for that domain name"), and normally use well-known port numbers (if you send a UDP packet to port 53, network firewalls will assume that it is a DNS query, and unlikely to cause congestion).

HTTP/3, like any other protocol running over QUIC, will look like a highly encrypted UDP protocol, and it **might** use the UDP port number that matches the TCP port number of the same protocol running over TCP, but that is not a requirement, so network operators often investigate UDP packets being sent to an unfamiliar port number, especially if that traffic does not seem to be simple query-response traffic.

If their investigation is not reassuring, they may block UDP packets being sent to an unfamiliar port number at an unfamiliar IP address, and even if they don’t block that traffic, they may rate-limit the traffic to prevent their network links being overwhelmed by unknown traffic that might not respond to congestion indications. So, on the Internet, HTTP applications that attempt to use HTTP/3 are prepared to fall back to HTTP/2 or even HTTP/1.1 over TCP, which is more reassuring for network operators.

#### 5.4.5.4 Adaptive Streaming clients operating on top of HTTP/3 capabilities

Adaptive streaming clients are implemented to overcome the TCP-related deficiencies that HTTP implementations were forced to deal with. Today’s adaptive streaming clients are typically not aware that they are operating on top of HTTP/3 and QUIC instead of HTTP/1.1 and TCP. In order for the browsers to establish an HTTP/3 connection, they try to initialize both a TCP and a QUIC connection, via “connection racing” techniques that are often inadequate to provide the optimal solution [x2]. Moreover, according to [x1][x2] in current browser implementations during a session there can be a switch to/from TCP and UDP, resulting in a non standard-compliant behaviour of using HTTP/3 over TCP (presumably in a manner similar to [x3]), having a negative impact in performance. A DASH client as documented in clause 13 of TS 26.512 [?] includes some typical functions that may be impacted by operation on top of HTTP/3, in particular:

- *Throughput estimation:* An estimate of the network throughput from the 5GMSd Application Server, which is typically computed as the object size divided by the download time where the download time is the time difference between the first and the last bytes received for that object. According to [DASH-QUIC], measuring the throughput of multiplexed audio and video streams over a single UDP socket results in additional response latency for the (much smaller) audio segments, whose individual throughput is not captured during the calculation of channel throughput. HTTP/3-specific algorithm selection methodologies exist that consider the underlying QUIC connection [x4] [x5].

- *Request Scheduling:* The adaptive streaming client schedules requests under the assumption of TCP operation. It typically operates for example audio and video on two HTTP sessions with separate sockets, and each of the sockets maintains its own independent socket buffer. By contrast, requests for audio and video are typically multiplexed onto a single HTTP/3 session on a single UDP socket with a single socket buffer. Therefore, the HTTP responses from both the streams interfere, and a higher response rate for one stream affects the queuing delay for the responses on the other stream.

Based on these operations, it is important that the adaptive streaming client:

1. Is aware that it is operating on top of HTTP/3.

2. Adapts its operation based on HTTP/3 properties

Details are for further study.

#### 5.4.5.5 5GMS Operation taking advantage of HTTP/3 capabilities

5G Media Streaming provides the ability to support regular OTT media streaming by providing additional and auxiliary information between the Media Session Handler and the 5GMSd AF. Supported functions in Rel-16 include telco CDN, network assistance and for example metrics reporting. It would be appropriate to adjust 5GMS function to HTTP/3 based delivery. As an example, certain DASH metrics are designed for TCP based streaming and would preferably be updated to account for HTTP/3 based delivery. Additionally, novel ABR algorithms designed for HTTP/3 delivery [x5] take in consideration QLOG metrics in their bitrate selection logic.

Editor’s Note: Identify the impact of including QLOG events on 5GMS metrics reporting.

Change #3

#### 5.4.5.8 Computational overhead of HTTP/3 clients

Adoption of HTTP/3 can have impact on the UEs due to the processing required for QUIC. In TCP most functions were executed in kernel space, which over the years have been heavily hardware-optimized to handle, while in QUIC most functions have moved to user space.

There are two identified sources for this computational overhead [x6]. First, the in-kernel UDP stack issues many packet reads; this is because each datagram arriving in the link layer is forwarded for processing to the transport layer - i.e., there are no offload mechanisms used like the sender-side Generic Segmentation Offload (GSO), or the generic receive offload mechanism (GRO) that the link layer module combines datagrams into a mega datagram before being forwarded to the transport layer. Secondly, there is increased overhead on processing packets (and as a result in generating responses) due to increased processing required on user-space cause by the first mentioned issue (i.e. all the packets need to be processed individually) combined with managing QUIC ACKs in user-space (instead of kernel-space as with TCP in HTTP/2).

These potential open issues should be taken in consideration because they can impact the QoE and increase the energy consumption of UEs.

END OF CHANGES